DOCUMENT RESUME

ED 164 551 TM 007 154

AUTHOR Comerford, James P.; Scheib, James E.
TITLE The Construction of a Machine-Scored Reading Inventory for Use In Secondary Compensatory Reading Programs, and an Investigation Into Its Sensitivity to Instruction.
PUB DATE Mar 78
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Criterion Referenced Tests; Elementary Secondary Education; Grade Equivalent Scores; *Junior High School Students; Low Ability Students; *Low Achievers; Reading Achievement; *Reading Level; *Reading Tests; *Test Construction; Test Reliability; Test Validity
IDENTIFIERS Elementary Secondary Education Act Title I; *Secondary Reading Assessment

ABSTRACT

As a supplement to standardized reading tests, the Philadelphia Public School System prepared three parallel sets of reading tests to meet the following criteria: (1) be suitable for grouping secondary pupils within Title I reading classes; (2) be group administered and machine scoreable; (3) be short enough to be administered in one or two regular classroom sessions; (4) have a range of reading levels from grades one to nine; (5) be of interest to junior high pupils; (6) have minimal costs; and (7) have parallel forms. The tests were assembled, pretested on a group of 400 secondary school pupils, revised, and readministered to about 800 pupils in grades 8 and 11. Reliability data are reported. Test difficulty was assessed by various readability formulas and validated by comparing test scores with teachers' judgments of student reading ability. The new test, the Secondary Reading Assessment (SRA) was correlated with the California Achievement Test. The SRA was judged to be useful over a wide range of reading ability with a minimum number of chance scores. (Author/CTM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

#16.22

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED164551

# THE CONSTRUCTION OF A MACHINE-SCORED READING INVENTORY FOR USE IN SECONDARY COMPENSATORY READING PROGRAMS, AND AN INVESTIGATION INTO ITS SENSITIVITY TO INSTRUCTION.

JAMES P. COMERFORD

JAMES E. SCHEIB

SCHOOL DISTRICT OF PHILADELPHIA

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James P. Comerford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the annual meeting of the American Educational Research Association, Ontario, Canada, March, 1978.

Printed in U.S.A.

TM007 154

INTRODUCTION

The use of norm referenced, in-level, standardized, achievement tests for pupil placement within Title I programs or for the evaluation of such programs is often less than desirable because of the difficulty level of the tests. Teachers sometimes complain that what is taught is not what is tested (Stuck, 1977), and that it is difficult to motivate students to attempt a difficult test. Scheib (1974), for example, has suggested that the reliability of in-level standardized test scores for low-achieving pupils is highly suspect. Alternatives to this problem include out-of-level testing, criterion referenced testing and matrix sampling. Problems associated with these procedures include large financial expenditures, test length and the suitability of the test content for the age group to whom the tests are administered. In elementary grades in addition to standardized tests, pupils are regularly administered an informal reading inventory to determine their instructional reading level. Such an inventory is hand-scored and administered to individuals or in small groups. In secondary schools this procedure is difficult because of the wide range of reading levels and because each teacher often teaches over 100 pupils.

As a supplement to standardized reading tests, the Philadelphia Public School System, Districts Two and Three, began a search for an inventory to meet the following criteria: (1) it must be suitable for grouping secondary pupils within Title I reading classes; (2) it must be group administered and machine-scoreable; (3) it must be short enough to be administered in one or two regular classroom sessions; (4) it must have a range of reading levels from grades one to nine; (5) the interest level of the test content must match that of junior high pupils; (6) the cost must be minimal; (7) it must have at least two

parallel forms.

Since no test could be found that met all seven criteria, it was decided to construct a machine-scored, multiple choice reading inventory consisting of three forms and covering nine reading levels.

DESCRIPTION OF THE INSTRUMENT

Over a period of two years, three parallel forms of the Secondary Reading Assessment (SRA) were developed. Each form (A,B, and C) has nine passages in ascending order of difficulty from reading levels one to nine printed in three booklets; each booklet contains 3 passages and 30 items. There are ten multiple-choice questions on each passage for a total of ninety questions in each form (3 booklets).

To match test difficulty with pupil achievement level, each pupil is administered either booklets 1 and 2 (items 1-60), or booklets 2 and 3 (items 31-90). The overlapping design allows the teacher to administer different levels within the same class. Because of the continuous numbering of items, only one answer sheet is needed. If, after the first administration, the teacher decides that a particular student should take either a higher or lower level, the student need only answer an additional 30 items (1-30 or 61-90) on the same answer sheet. All answer sheets are then divided into two groups: items 1-60 and items 31-90. These are scanned and the results processed by computer to produce individual scores and item results, as well as summaries for each classroom, grade and school.

RELIABILITY

In May, 1975 the inventory was administered to approximately 400 secondary pupils from 5 schools. Grades 8 and 11 were chosen from the sample since

they represent the middle grades in junior and senior high.

Teachers were asked to identify pupils reading between levels 1-3, 4-6 and 7-9. These pupils were then given the appropriate booklets of the inventory. Each pupil was administered all three forms of the test.

The general findings of this initial try-out are presented in Table I. The results indicated that in general the test is properly leveled. Compared to previous estimates of pupils' reading ability, the pupils in the sample performed as expected. That is, within each leveling group (1-3, 4-6, 7-9) relatively few pupils scored at or below the guessing level, and few pupils scored above the mastery level. Items were edited on the basis of the item analysis.

In January, 1976, the newly edited inventory was administered to a sample of about 800 secondary pupils (grades 8 and 11). As in the first try-out, each pupil was administered a series of three levels on each of the three forms. Unlike the first try-out, no attempt was made to test pupils by previous reading levels. In fact, the design specified that within each school pupils would be randomly assigned to the booklet levels of the test. This design enables one to test not only the reliability among alternate forms, and the appropriateness of the items, but also the difference in difficulty between booklet levels of the test.

As can be seen from Table II, the difference in mean scores between booklet levels is what one would expect. The average score for booklet levels 1-3 is higher than 4-6, and 4-6 is higher than booklet levels 7-9. The difference between average scores across forms is small. In all but one case, the

difference in average scores is less than two items. To measure the reliability among the three forms, Pearson product moment correlations, within group correlations, Livingston's criterion referenced reliability estimates, Livingston's reliability of classification, and Kelley's triad reliabilities were computed (Tables III and IV). Although the five indices were found to vary considerably, the instrument was shown to have demonstrated adequate reliability for most purposes and very high estimates of reliability for others.

Test statistics generated from the two pilot administrations indicate that among the three forms of the test there is an acceptably high reliability of measurement, reliability of classification, and internal consistency. The three forms will be equated so as to produce a table of equivalent scores. Completed portions show that this table will not differ substantially from the raw scores.

VALIDITY

An important question to be discussed is the validity of the test. That is, does the test measure what we intend it to measure? Do the test scores in fact reflect the reading levels of the pupils? A review of the literature on Reading Inventories reveals some interesting findings. Kender (1966) found that different reading inventories yeild different results for the same students. The reason for this is quite obvious: there is no single definition of what constitutes a particular reading level.

To read at a particular reading level means to be able to read and comprehend the words, facts, ideas, inferences of a passage with a certain level of competency. It is crucial, therefore, to identify the readability level of a

passage. Different readability formulas, however, yield different results for the same passage. Some formulas stress sentence length, others stress the number of polysyllabic words, others stress sentence complexity and so on (Klare, 1974).

The measure of a pupil's reading level is also influenced by the subject matter of the passage. Using the cloze procedure to measure pupils' ability to read various types of materials, Hansen and Hesse (1974) found different achievement levels for each of ten domains of materials (school-related, recreational, occupational, etc.)

One solution to this problem is to draw passages from reading materials in which the student is going to be instructed. The logic is that the test results will be used primarily for diagnosis and instruction (Kender, 1966).

McCracken (1963) followed a similar procedure in developing his Standard Reading Inventory. He chose materials from three basal series. He hypothesized that the passages were leveled correctly if a sample of children (664) found succeeding difficulty in reading the passages. He further hypothesized that the test was properly leveled if children found succeeding difficulty in answering the questions on each level. As a further check on the readability of the passages, he correlated the levels with the opinion of twenty-five reading experts.

In developing the SRA it was hypothesized that the passages were leveled correctly if: (1) a sample of pupils tested found succeeding difficulty in reading the passages, (2) if pupils found succeeding difficulty in answering questions on each level, and (3) if the leveling of passages substantially agreed with teacher judgement.

Therefore, after editing the reading passages, various readability formulas were applied to them (Smog, Fry, Cloze). To further assess the readability levels of the passages, an oral reading design was constructed in which children whose reading levels were previously determined by the teacher were asked to read orally the new reading passages. After the second pilot administration of the test, teachers were asked to assess individual pupils' test results in relation to their previous assessment of the pupils' reading ability. The results of applying readability formulas, an oral reading design, teacher judgement and test score results indicates that the passages are properly leveled.

CONSTRUCTION OF READING LEVEL SCORES

The final step in developing the Secondary Reading Assessment was the conversion of raw-scores into reading level scores. The method outlined here is the total raw-score method, which is an adaptation of the method used by McCraken, Johnson and Kress (1965) and others.

This method uses a raw-score interpretation for each individual subtest. In some cases this is done because no overall reading level is predicted and because all pupils do not take the entire battery of subtests. The Standard Reading Inventory, for example, uses the following criteria: 90 to 100% correct equals an independent level, 60 to 80% equals the instructional level, and 0 to 50% equals the frustration level on each subtest. The setting of cut-off scores is arbitrary, but provides objectivity in scoring all pupils. This method assumes that pupils will score lower on each successive level of difficulty. An empirical test of this

hypothesis, using a local unpublished inventory, revealed that too often this was not the case, even when the criteria for each subtest were adjusted for overall difficulty. The reasons for this lack of consistency for individuals is due to a number of factors: guessing, familiarity with passage subject matter, item difficulties, random error, and different abilities to handle factual, inferential, and vocabulary questions.

The method used to set the reading level scores for the SRA was to set the minimum and maximum interpretable scores and then to divide the remaining scores into intervals to represent the reading levels. The average theoretical guessing level score of 15 out of 60 items for four-choice multiple choice questions can be obtained on the average without reading the questions, but only marking the items by random selection of answers. An unpublished study conducted by the writers using a similar reading inventory has shown empirical guessing levels to be somewhat above the theoretical guessing score. It was decided not to interpret any scores below 20 as having shown enough evidence of reading ability to assign a reading level score.

A score of 60 items correct (100%) would likely be attained by readers who were in fact higher in their reading ability than the highest level tested, therefore, the upper most category was made indeterminate in its ceiling (e.g. 6th reading level or above, 9th reading level or above). The forty interpretable scores between 21 and 60 were divided into 6 roughly equal intervals and assigned reading levels of 1,2,3,4,5,6-or-above (for tests where the first two booklets were administered) and 4-or-below, 5,6,7,8,9-or-above (for tests where the middle and last booklets were administered).

A final aspect of test development to be considered is the motivation of the pupils taking the test. Pupil motivation is affected by a number of factors, including the interest level of the reading passages, test difficulty and attitude of the teacher administering the test.

Part II of this paper will discuss the difficulty levels of the test in relation to a standardized test. It will be shown that the difficulty levels on the SRA are appropriate for the reading levels of the pupils tested. Although no formal measure of pupil attitude toward the test was taken, the authors personally observed the administration of both pilot testings. In those settings, pupils displayed positive attitudes toward the test. Few pupils showed frustration with the testing situation. Comments solicited from pupils after the testing session indicated that they found the reading passages interesting and enjoyable.

The attitude of teachers administering the test has been a prime concern of the test developers. If teachers perceive a test as being valuable and useful for their pupils, it is likely that they will communicate that positive attitude to their pupils. To facilitate this, the following procedure was used: (1) the lead reading teacher in each school was asked to submit reading passages and questions for inclusion in the test; (2) at each phase of test development, a progress report was given to these teachers; (3) copies of each try-out edition were given to teachers for their criticism; (4) at each discussion session of the instrument, the attitude conveyed was, "This is your test. It was developed at your request, with your input, for your pupils." (5) finally, the test was to be used by the classroom teacher for initial placement of pupils.

SENSITIVITY TO INSTRUCTION

Part II of this paper concerns the relative sensitivity to instruction of the SRA. It has been suggested by teachers and administrators that pupils often show little or no progress on standardized tests because the tests are not sensitive to instruction. This can be due to a number of factors. Sensitivity to instruction is a term sometimes used by test developers to refer to content or curricular validity. That is, it can be demonstrated empirically that a test is sensitive to instruction when, after adequate instruction, pupils will show significant gains over their pre-test scores.

Thus, in a particular instructional setting, the sensitivity to instruction of test items is related to (1) the congruence of the items to curricular content, (2) the difficulty levels of the items in relation to the population tested, (3) the motivation of the examinees, and (4) the strength of the particular instruction.

Tennenbaum and Miller (1977) developed a methodology to measure the congruence "between the knowledge and skills tested by a norm-referenced achievement test (MAT '70) and the content in the curricula specified in six, exemplary, compensatory education programs." For reading comprehension, they compared the difficulty levels of the passages and the type of item (main idea, literal, inferential) with those same elements of the reading curriculum. Preliminary findings indicated that for grade eight pupils, only 2 of 50 Word Knowledge items and only 3 of 45 Reading Comprehension items were found to be congruent with the curriculum.

Congruence between items and curricular content was a particular concern in the contruction of the SRA. Careful attention was given to the readability

level of the selections, proper passage length, and types of items constructed. It was, therefore, judged that the items were in fact congruent with the reading curriculum. Any direct measure of sensitivity to instruction via pre-tests and post-tests would only be some combination of the test's appropriateness and the strength of the instruction.

It was, therefore, decided to examine the difficulty level of the SRA compared to that of a standardized reading test as an indirect measure of sensitivity to instruction. When the difficulty of a test is perceived to be too high, test frustration results in unequal effort on the test over different testings. An indicator of this is the percent of scores at or below the theoretical guessing level.

Ideally, no pupil should score below that point at which he could score by randomly choosing responses without reading the questions. The percent of guessing level scores for each test was computed and compared as an index of difficulty and appropriateness of the tests for the population involved.

A test may also be inappropriate because it is too easy. In this case, a student may not be able to demonstrate his full ability in the particular skills being tested. As an index of the ceiling effect of the test, the percent of pupils who correctly answered ninety percent or more of the items was calculated.

Mean, and median scores as well as Pearson product moment correlations were also calculated. In addition, the distribution of guessing level scores was also examined.

The sample population consisted of 704 pupils in grades seven to nine who were enrolled in Title I reading classes. The SRA, Form A was compared to

level four of the California Achievement Tests, 1970 edition, subtest Total Reading.

A guessing level score was defined as the total number of items divided by the number of choices per item. On the SRA there are 60 items with four choices per item. A guessing level score is, therefore, from 1 to 15 items correct. (Zero scores were deleted).

There are 85 items on the Total Reading subtest of the CAT. Seventy-six items have four choices each, and 9 items have five choices each. A guessing level score is, therefore, from 1 to 21 items correct.

The results indicated that 15.34% scored at or below the guessing level on the CAT, while only 2.13% scored at or below the guessing level on the SRA (Table VI). Of the same sample, 3.12% correctly answered ninety percent or more of the items on the SRA, while 0.28% correctly answered ninety percent or more on the CAT.

Thus the difficulty level of the items on the SRA seems more appropriate for the reading levels of the population tested. One might reasonably expect, therefore, that the test would be appropriate for placing pupils within reading programs and that the SRA would be sensitive to changes in pupil performance due to adequate instruction.

Correlation indices were computed between the SRA and the CAT in three ways (Table VII): (1) reading level scores on each test for the whole group, (2) SRA raw scores and CAT scale scores for those who took SRA, level 1 (items 1-60) and (3) those who took SRA, level II (items 31-90). The results indicate a correlation between the two tests of 0.639. SRA level II and the CAT were correlated higher (.692) than SRA, level I and CAT (.487). This is attributed

to the fact that scores of low scoring pupils were considered to be less reliable than those of higher scoring pupils.

Comparisons between the mean raw score on the SRA, level I and the CAT and SRA, level II, and the CAT revealed the following: (1) Pupils who took level I of the SRA obtained an average raw score of 37.8 which is more than 50% correct. On the CAT these same pupils obtained an average scale score of 400 (equal to a raw score of 27) which is only 32% correct and only 6 items above a guessing level score. (2) The interpreted reading level score of pupils who took the SRA, level I (G.E. of 3) was lower than the interpreted reading level score on the CAT (G.E. of 4.4). (3) The average raw score of pupils who took SRA, level II and CAT was well above the guessing level on both the SRA (35.7 of 60 items) and the CAT (34 of 85 items). (4) The interpreted reading level scores of pupils who took SRA, level II and the CAT were similar (G.E. of 6 on the SRA and a G.E. of 5.6 on the CAT). See Table VIII.

Thus, while lower scoring pupils obtained an average percent correct on the SRA which was higher than their percent correct on the CAT, their interpreted score was lower on the SRA. This difference highlights the fact that a grade equivalent or a reading level score is an interpreted score based upon judgement and will, therefore, vary somewhat between tests. Furthermore, this difference seems greater at the lower end of the raw score scale. It should be noted that on the CAT, level 4, a guessing level score is equivalent to a reading level score of 3.2. An examination of the distribution of scores on the CAT and SRA revealed that pupils who scored below the guessing level on the CAT ranged from a reading level score of 0.6 to 5.0

on the SRA. Developing a test to meet the difficulty level of the population tested does not, therefore, guarantee that the average interpretated test score will be higher than on an in-level standardized publisher's achievement test.

## SUMMARY AND CONCLUSIONS

The SRA was constructed as a supplement to standardized achievement tests. The purpose of the instrument is to provide data for placing pupils within Title I reading classes. It is group administered and machine-scoreable. The overlapping test design enables the teacher to test pupils whose reading ability ranges between levels one and nine, yet the interest level of all passages is aimed at junior high. The three parallel forms allow for annual testing of the same students without overusing a single form.

A comparison of the SRA and a norm-referenced standardized achievement test highlights the different purposes for which each was constructed. As a norm-referenced test, the CAT was constructed to yield maximum differentiation among the population tested, while the SRA, which is identified closely to a criterion referenced test, is designed to identify each examinee's reading proficiency according to pre-defined criteria.

Standardized tests still play a vital role in a school district's measurement program. They serve to assess the overall reading progress of the district's pupils; they help identify both the academically talented and those who need special instruction. In the sample studied, all of the pupils who scored below the guessing level on the CAT also scored at least two years below grade level on the SRA. Thus standardized tests continue to be a

practical and appropriate method for identifying pupils for compensatory education.

An examination of the percent of guessing level scores on the two tests suggests, however, that other instruments more specifically geared to the reading levels of the population tested might also be useful in grouping pupils within compensatory education programs, and in measuring their progress in that program.

SUGGESTIONS FOR FURTHER STUDY

To further assess the possible differences between an in-level, norm-referenced achievement test and a criterion referenced instrument, change scores on both the CAT and the SRA should be studied and analyzed. One might hypothesize that more positive changes would occur on the SRA than the CAT because of the difference in difficulty levels.

Before change scores on different forms of the SRA are compared, an equating study must be made on the three forms of the SRA.

TABLE I

TEST STATISTICS FROM FIRST

PILOT ADMINISTRATION - MAY, 1975

| FORM | TEST LEVEL | N | # ITEMS* | MEAN | MEAN/# OF ITEMS | S.D. | S.E.M. | KR20 |
|---|---|---|---|---|---|---|---|---|
| A | 1-3 | 107 | 30 | 15.88 | 0.53 | 8.09 | 2.791 | 0.869 |
| B | 1-3 | 114 | 31 | 14.17 | 0.46 | 8.32 | 3.060 | 0.861 |
| C | 1-3 | 110 | 30 | 18.48 | 0.62 | 8.42 | 2.679 | 0.893 |
| A | 4-6 | 122 | 40 | 15.92 | 0.40 | 9.65 | 3.441 | 0.872 |
| B | 4-6 | 121 | 32 | 14.68 | 0.46 | 7.09 | 3.067 | 0.811 |
| C | 4-6 | 122 | 30 | 15.48 | 0.52 | 8.36 | 2.898 | 0.882 |
| A | 7-9 | 166 | 46 | 26.57 | 0.58 | 10.38 | 3.470 | 0.920 |
| B | 7-9 | 160 | 31 | 18.97 | 0.61 | 6.43 | 2.793 | 0.810 |
| C | 7-9 | 161 | 30 | 16.11 | 0.54 | 5.45 | 2.890 | 0.700 |

*Where possible, extra items were included so that a choice could be made of the best ten items.

TABLE II

TEST STATISTICS FROM SECOND

PILOT ADMINISTRATION - JAN., 1976

| FORM | TEST LEVEL | N* | MEAN | S.D. | S.E.M. | KR20 |
|---|---|---|---|---|---|---|
| A | 1-3 | 242 | 25.68 | 4.62 | 1.710 | 0.853 |
| B | 1-3 | 242 | 24.36 | 4.89 | 1.884 | 0.866 |
| C | 1-3 | 242 | 25.65 | 4.30 | 1.691 | 0.849 |
| A | 4-6 | 177 | 20.72 | 6.05 | 2.078 | 0.880 |
| B | 4-6 | 177 | 19.05 | 7.02 | 2.164 | 0.928 |
| C | 4-6 | 177 | 20.89 | 6.27 | 2.097 | 0.890 |
| A | 7-9 | 234 | 18.67 | 7.05 | 2.146 | 0.905 |
| B | 7-9 | 234 | 18.34 | 6.46 | 2.178 | 0.871 |
| C | 7-9 | 234 | 16.51 | 7.36 | 2.229 | 0.904 |

*Test results are reported for those pupils who completed all three forms of the test.

SECONDARY READING ASSESSMENT

KELLEY TRIAD RELIABILITY ESTIMATES (KT)

*LIVINGSTON CRITERION REFERENCED RELIABILITY ESTIMATES (k)

*LIVINGSTON RELIABILITY OF CLASSIFICATION (c)

| TEST | KT AA | k A-B | c A-B | KT BB | k B-C | c B-C | KT CC | k A-C | c A-C |
|---|---|---|---|---|---|---|---|---|---|
| 1-3 | .699 | .826 | 89.67% | .550 | .871 | 90.90% | .524 | .896 | 93.80 |
| KR20 | .853 | | | .866 | | | .849 | | |
| 4-6 | .699 | .746 | 78.53 | .785 | .848 | 82.48 | .737 | .776 | 81.92 |
| KR20 | .880 | | | .928 | | | .890 | | |
| 7-9 | .834 | .882 | 82.40 | .754 | .817 | 78.63 | .844 | .824 | 80.77 |
| KR20 | .905 | | | .871 | | | .904 | | |

*A raw-score of 17 was chosen as a criterion score for all levels and all forms.

TABLE IV

SECONDARY READING ASSESSMENT

PEARSON PRODUCT MOMENT CORRELATIONS (r)

WITHIN GROUPS CORRELATIONS (w)

| TEST | A - B | | B - C | | A - C | |
|---|---|---|---|---|---|---|
| | r | w | r | w | r | w |
| 1-3 | .585 | .620 | .518 | .537 | .574 | .605 |
| 4-6 | .734 | .741 | .745 | .761 | .690 | .718 |
| 7-9 | .811 | .793 | .825 | .798 | .833 | .839 |

TABLE V

RAW SCORE TO READING LEVEL EQUIVALENTS
FOR ITEMS 1-60 AND ITEMS 31-90

| | READING LEVEL | |
|---|---|---|
| R.S. | ITEMS 1-60 | ITEMS 31-90 |
| 0-20 | GUESSING | GUESSING |
| 21-27 | 1 | 4 |
| 28-34 | 2 | 5 |
| 35-41 | 3 | 6 |
| 42-48 | 4 | 7 |
| 49-54 | 5 | 8 |
| 55-60 | 6 or above | 9 or above |

TABLE VI

PERCENT OF PUPILS SCORING BELOW THE GUESSING LEVEL AND ABOVE THE CEILING LEVEL ON THE SRA AND CAT

| N | BELOW GUESSING LEVEL | | | |
|---|---|---|---|---|
| | SRA | | CAT | |
| | N | % | N | % |
| 704 | 15 | 2.13 | 108 | 15.34 |

| N | ABOVE GUESSING LEVEL | | | |
|---|---|---|---|---|
| | SRA | | CAT | |
| | N | % | N | % |
| 704 | 22 | 3.12 | 2 | 0.28 |

TABLE VII

PEARSON PRODUCT MOMENT CORRELATION INDICES BETWEEN

SRA AND CAT

| | | GRADE EQUIVALENT | ACHIEVEMENT DEVELOPMENT SCALE SCORE |
|---|---|---|---|
| SRA | READING LEVEL SCORE | 0.639 (N = 704) | |
| | RAW SCORE (LEVEL I) | | 0.487 (N = 480) |
| | RAW SCORE (LEVEL II) | | 0.692 (N = 224) |

TABLE VIII

COMPARISON OF AVERAGE SCORES ON SRA AND CAT

| | | MEAN SCORE | | | |
|---|---|---|---|---|---|
| SAMPLE | N | SRA | | CAT | |
| | | Raw Score | Reading Level | Raw Score | Grade Equivalent |
| SRA (Level I) | 480 | 37.8 | 3 | 27 | 4.4 |
| SRA (Level II) | 225 | 35.7 | 6 | 34 | 5.6 |

## BIBLIOGRAPHY


Hansen, Lee H., and Hesse, Karl D. A Criterion-Referenced Assessment of Reading Literacy Using the Cloze Procedure. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Ill., April, 1974.

Johnson, Marjorie S., and Kress, Roy A. Informal Reading Inventories. International Reading Association, Newark, Del., 1965.

Kender, Joseph P. An Analysis of Factors Associated with Informal Reading Tests at the Eighth Grade Level. Unpublished dissertation, University of Pennsylvania, Phila., 1966.

Klare, George R., Assessing Readability. Reading Research Quarterly, 1974-1975, 10, 62-102.

McCracken, Robert A. Standard Reading Inventory. Pioneer Printing Co., Bellingham, Washington, 1963.

Scheib, James E. A Descriptive Study of Longitudinal Growth in Reading as Measured by Standardized Tests in an Urban Population. December, 1974 (Unpublished paper, University of Pennsylvania).

Stuck, Gary B. and Wyne, Marvin D. Teachers' and Pupils' Opinions Relative to the Major Issues in Collecting and Using Pupil Test Data. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.

Tenenbaum, Arlene B. and Miller, Christine A. The Use of Congruence Between the Items in a Norm-Referenced Test and the Content in Compensatory Education Curricula in the Evaluation of Achievement Gains. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.